

Durham Research Online

Deposited in DRO:

17 August 2020

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Bartholo, T.L. and Campelo Koslinski, M. and da Costa, M. and Tymms, P. and Merrell, C. and Mendonça Barcellos, T (2020) 'The use of cognitive instruments for research in early childhood education : constraints and possibilities in the Brazilian context.', *Pro-Posições*, 31 . e20180036.

Further information on publisher's website:

<http://dx.doi.org/10.1590/1980-6248-2018-0036>

Publisher's copyright statement:

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

**The use of cognitive instruments for research in
early childhood education: constraints and possibilities in the
Brazilian context**^{1 2 3 4}

***O uso de instrumentos cognitivos para pesquisa na educação
infantil: limites e possibilidades no contexto brasileiro***

Tiago Lisboa Bartholo ⁽ⁱ⁾

Mariane Campelo Koslinski ⁽ⁱⁱ⁾

Márcio da Costa ⁽ⁱⁱⁱ⁾

Peter Tymms ^(iv)

Christine Merrell ^(v)

Thaís Mendonça Barcellos ^(vi)

(i) Federal University of Rio de Janeiro – UFRJ, Rio de Janeiro, RJ, Brazil. <https://orcid.org/0000-0002-2400-8707>, tiagobartholo@gmail.com

(ii) Federal University of Rio de Janeiro – UFRJ, Rio de Janeiro, RJ, Brazil. <https://orcid.org/0000-0002-9644-5041>, mckoslinski@gmail.com

(iii) Federal University of Rio de Janeiro – UFRJ, Rio de Janeiro, RJ, Brazil. <https://orcid.org/0000-0002-5235-9782>, marcioc@pobox.com

(iv) Durham University – DU, Durham, DUR, United Kingdom. <https://orcid.org/0000-0002-7170-2566>, p.b.tymms@dur.ac.uk

(v) Durham University – DU, Durham, DUR, United Kingdom. <https://orcid.org/0000-0001-5757-6252>, christine.merrell@durham.ac.uk

(vi) IDados, Rio de Janeiro, RJ, Brazil. <https://orcid.org/0000-0002-7827-7400>, thais@idados.id

¹ Apoio: Inter-American Development Bank (IADB)
Instituto Alfa e Beto (IAB)

Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ)

² References correction and bibliographic normalization services: Andressa Picosque (Tikinet) – revisao@tikinet.com.br

³ Copy Editor: José Pereira de Queiroz – ze.pereira.queiroz@gmail.com

⁴ Responsible Editor: Profa. Dra. Helena Maria Sant'Ana Sampaio Andery - <https://orcid.org/0000-0002-1759-4875>, hsampaio@unicamp.br

Abstract: This paper discusses the adaptation of the iPIPS assessment for use in research with Brazilian children between the ages of four and seven. It debates the importance of having a baseline measure to assess early childhood education policy as well as the advantages of collecting high-quality information about children's development. Not knowing how children are progressing could harm disadvantaged children and increase school inequality. The data used in the analysis was based on the pre-test of the iPIPS 2016 for mathematics and language in a total of 560 cases collected in three Brazilian cities. The preliminary analyses indicate that the items of both tests present adequate behavior, suggesting the theoretical adequacy of the items and a good adaptation and application protocol.

Keywords: monitoring, childhood education, assessment, literacy

Resumo: O artigo discute a adaptação do instrumento Performance Indicator for Primary Schools (iPIPS) para uso em pesquisa no contexto brasileiro com crianças entre 4 e 7 anos. Debate a necessidade de construir medidas diagnósticas (linha de base) para avaliar o impacto das políticas da educação infantil. Desconhecer como as crianças se desenvolvem ao longo dos primeiros anos da escola pode prejudicar mais fortemente os alunos em situação de vulnerabilidade e potencialmente aumentar a desigualdade educacional. Os dados utilizados são do pré-teste do iPIPS realizado em 2016 como instrumento para medir linguagem e matemática com 560 crianças em três cidades brasileiras. Os resultados preliminares sugerem que o comportamento dos itens ocorreu da forma esperada, indicando adequação teórica. A escala do teste é adequada para medir o aprendizado o conhecimento de crianças entre 4 e 7 anos.

Palavras-chave: monitoramento, educação infantil, avaliação, alfabetização

Introduction

It is widely recognized that children's early development and their progress during the first years of school are crucial for their later success (Peisner-Feinberg et al., 2001; Sammons et al., 2008; Sylva, Melhuish, & Sammons, 2010; Sylva et al., 2006; Tymms, Jones, Albane, & Henderson, 2009). Then, measuring and monitoring children's evolution in this key stage of their life should be a policymakers' concern. Evidence to guide educational policy is important and this begins with understanding what children know and can do when they start school in their own country. This understanding sets out a Learning Path to inform curriculum

development and framework against which the impact of interventions may be evaluated. An understanding of how the development of children in one country is comparable to other countries at the start of school life can provide policymakers with information to help them evaluate the effectiveness of both preschool and educational policies.

In order to gain an understanding of young children's development, it is important to have high-quality instruments with the capability of measuring children's baseline at the start of school and their continuous learning during the first years in school. Another focus, as alluded to in the previous paragraph, can be a comparable measure across countries. In this sense, the iPIPS instrument (International Performance Indicators in Primary Schools; www.ipips.org) is an international assessment of children starting school which can also be repeated at the end of the first school year. iPIPS has the potential to fill an important gap in providing comparative information about children's progress in the first year of school around the world.

This paper discusses the adaptation of the iPIPS assessment for use with Brazilian children between the ages of four and seven. It debates the importance of having a baseline measure to assess early childhood education policy as well as the advantages of collecting high-quality information about children's development and skills on entry to school with the ability to predict later literacy and mathematics outcomes (Jordan, Glutting, & Ramineni, 2010; Schneider et al., 2017; Tymms, Merrell, & Henderson, 1997; Tymms, Merrell, & Jones, 2004).

In Brazil, in the wake of the inclusion of early childhood education in the system of basic education, proclaimed in the 1988 Constitution (Brasil, 1988) and consolidated by the 1996 Law no. 9,394, *Lei de Diretrizes e Bases da Educação* [Education Guidelines and Bases Law], the coverage at this educational level has achieved considerable advances. The growth in the number of enrolments is clear from the time series of the *Censo Escolar* [School Census]. In the case of children aged 0–3 years old, only 13.8% of the population were attended in 2001, rising to 30.4% in 2015, according to the *Observatório do PNE* [National Education Plan Observatory]. Regarding the age range 4–5, 66.4% were attended in 2001, reaching 90.5% in 2015 (Observatório do PNE, n.d.).

However, what has been observed is an expansion without suitable planning. This situation can generate low-quality services, including institutions maintained precariously, which

can be harmful to child development (Campos, 2010; Rosenberg, 1999a, 1999b)⁵. In Brazil, the monitoring of education's quality is mandated by law on the *Plano Nacional de Educação* – PNE [National Education Plan]. However, in what concerns early childhood education, the National Education Plan only sets targets related to access to early childhood education (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2014) and the proposal for the National Assessment of Early Childhood Education is still in elaboration. Moreover, there is a lack of studies to evidence the requirements of the law or to observe school characteristics and policies associated to child development. At the time of writing, there are no high-quality instruments in widespread use across the country to assess child development in preschool and, therefore, it is not possible to measure the impact of education policies on children's educational trajectory.

This paper is organized in three parts. The first part discusses evidences of early childhood education impact on educational trajectories. The second debates the importance of having a baseline and presents the characteristics and potential of iPIPS to provide information for policy evaluation and international comparisons. The third describes the steps to adapt iPIPS assessment scale to Portuguese (language and mathematics) and presents the main findings of the pre-test conducted in 2016 with the adapted instrument. A total of 560 children were assessed in three Brazilian cities in public and private schools. Preliminary analyses indicate that the items of both tests (language and mathematics) are psychometrically sound, suggesting the theoretical congruence of the items and a good adaptation and application protocol.

Assessing the impact of preschool: possibilities, constraints and evidence from the Brazilian context

Children's cognitive attainment and lifelong outcomes of social mobility have long been assumed to be the result of school enrolment and regular attendance. Early years education and readiness to start school have positive effects on a range of child outcomes that contribute to

⁵ The concept of quality of early childhood education that guide Campos's (2010) report is based on the *Parâmetros Nacionais de Qualidade para a Educação Infantil* [National Parameters for the Quality of Early Childhood Education] (Brasil, 2006). Their definition considers "the children, women, and families' rights, consolidated at the legislation the diverse necessities of small children, considering their age group, life conditions of their family, the cultural characteristics of their environment and their individuality... accumulated scientific knowledge within the fields of education, psychology, social sciences, health, etc." (Campos, 2010, p. 30).

life-long events. Studies conducted in diverse contexts indicate that attending preschool is an effective policy in promoting greater equality of educational opportunities. It contributes to (future) learning and a more fluid, more long-lasting school trajectory, especially among deprived children—from low socio-economic backgrounds (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Peisner-Feinberg et al., 2001; Sammons et al., 2008; Schweinhart, 2002; Sylva et al., 2006). This section discusses the current potentiality and drawbacks of the large-scale assessment systems available in the Brazilian context to develop an analysis of this kind.

International longitudinal studies confirm that children that have had the opportunity to attend good-quality early childhood education environments present better cognitive and social skills during their elementary school. Two studies in particular present more robust evidence. The first was conducted in the USA (Peisner-Feinberg et al., 2000), and the second in the UK (Sammons et al., 2008; Sylva et al., 2003; Sylva et al., 2006; Sylva et al., 2010). The investigations estimated the impact of early childhood education school quality (measured on the ECERS-R scale – Early Childhood Environment Rating Scale: Revised Edition and on the ECERS-E – Early Childhood Rating Scale - Extension) on children’s school trajectories.

Heckman and colleagues argue that the so-called “cognitive” and “non-cognitive” skills developed in early interventions in children’s education are not only important for their development later in life, but also that these initial investments are significant for subsequent development (Heckman, 2000, 2008). Cunha, Heckman, Lochner, and Masterov (2006), state that the mastery of skills at a certain stage of the life cycle allows greater skill development in later phases (a phenomenon called “self-productivity”). Moreover, according to the authors, there is a complementarity of the investments; early investments facilitate the productivity of subsequent investments, that is, the former expand the latter’s effectiveness.

Both of the aforementioned studies provide strong evidence to suggest that: (i) attending preschool increases chances of future success for students in the schooling process; (ii) attending a high-quality preschool increases subsequent learning gains (when compared to the conditions of not attending preschool at all or of attending one of inferior quality), with a special highlight on deprived children (poor families and/or of a low socio-economic level).

In Brazil, there are very few studies on the impact of early childhood education on students’ learning and school trajectory. However, Damiani, Dumith, Horta, and Gigante (2011)

observed that students who attended preschool tended to have longer school trajectories. Another study, conducted in 2010 in three Brazilian state capitals, investigated the impact of preschool quality (measured on the ECERS-R scale) on language development measured in the second year of primary school (Brazilian Standardized Test “*Provinha Brasil*”). The data indicated that children who attend good-quality institutions presented better performance in the test than children who had not frequented preschool at all or had gone to substandard institutions (Campos et al., 2011). Nevertheless, the few studies carried out in the Brazilian context presented limits to tracing causal relations between preschool quality and students’ learning or their future trajectories. Most of them were not longitudinal, did not measure children’s developmental levels at the start of preschool, and only produced measures about school inputs, such as infrastructure and pedagogical materials and/or of students’ later outcome. Without a baseline measure, it was not possible to assess the progress made by children between the start of preschool and later outcomes.

iPIPS Brazil can make a significant contribution to policymakers and practitioners by associating measures about children’s development (cognitive and motor skills) with preschool/primary education policies and teachers’ practice. Such a contribution becomes more pressing in the face of the almost complete universalization of attendance in elementary school and the shift in the discussion about access to school for effective learning or teaching quality provided by schools (Bonamino & Oliveira, 2013).

In the last 20 years, Brazil has created broad educational assessment systems such as the *Sistema de Avaliação da Educação Básica* (SAEB) [Basic Education Assessment System], currently composed by the *Avaliação Nacional da Educação Básica* (ANEB) [National Basic Education Assessment], the *Avaliação Nacional do Rendimento Escolar* (ANRESC/Prova Brasil) [National School Performance Assessment], and the *Avaliação Nacional da Alfabetização* (ANA) [National Literacy Assessment], as well as various municipal and state systems with their own characteristics. All these assessment systems bear the characteristic of monitoring the educational situation by means of a sequence of items employing a comparable metric over time and presenting the evolutionary trends in the quality of Brazilian education. For this reason, it can be affirmed that the demand for information about the state and the evolution of the quality of Brazilian education has been reasonably satisfied by the assessment structure in place.

Nevertheless, the data are cross-sectional and insufficient to draw causal inferences regarding the impact of educational policies and school practices or school factors that influence pupils' learning. Franco (2001) argues that SAEB data has limitations in this respect, as proficiency is a measurement at only one moment, and, therefore, does not express the children's learning over the years (or even at a particular school stage). Longitudinal studies are considered the "gold standard", since the measurement of previous proficiency is used as a control, thus enabling better estimation of the effects of school in its multiple facets.

In Brazil, studies with more robust designs that allow estimation of the effects of specific educational practices or policies are scarce. The last longitudinal study with a probabilistic sample, the *Estudo Longitudinal da Geração Escolar 2005*—GERES [Longitudinal Study of the 2005 School Cohort] project, conducted in Brazilian state capitals, took place almost 10 years ago. It had three key objectives: (i) identifying school characteristics that maximize student learning and minimize the impact of social background on their learning; (ii) identifying school factors that decrease the probability of grade retention, and (iii) identifying those school characteristics that reduce the probability of absenteeism.

The project presented a design adequate for the objectives proposed, with different measurement waves of learning, which started in the former 1st grade (currently the 2nd year of elementary school) and ended in the 4th grade (currently the 5th year of elementary school). GERES created specific tests in mathematics and language, equalized with the SAEB scale. Nevertheless, as the study made its first data collection in the 1st grade of elementary school, the design did not allow making inferences regarding the impacts of early childhood education and the then so-called literacy class on the trajectory of students' learning.

The iPIPS: structure and potential use

PIPS, *Performance Indicators in Primary Schools*, has been developed and used for more than 20 years in the UK and other countries (Tymms, 1999a). The aim was to create a baseline measurement for use with children at the start of formal schooling. In England, where PIPS was first developed, this meant that the assessment was used with children aged 4–5 (the *Reception Year*). The assessment is underpinned by developmental and pedagogic roots with the

aim of providing information to help teachers target their practice effectively. In 2014, around 30% of the English state schools voluntarily used the assessment as a diagnostic pedagogic tool.

The PIPS assessment is designed, amongst other things, to track children's progress in reading and mathematics. It was originally created to act as a baseline for children starting school in England so that progress could be assessed, at a later stage, using value-added models (Tymms, 1999a). The development work therefore focused on the best predictors of success in literacy and numeracy. There was no shortage of information on the early indicators of literacy, but there were gaps in the literature on early indicators of numeracy, and work was devoted to filling this gap (Tymms 1999b).

The project evolved and, at teachers' instigation, the children were reassessed after just one year at school and the assessment was extended to include reading and numeracy *per se*; not simply the predictors. As the project expanded internationally, adjustments were made for each country, but the essence of the assessment remained, allowing the progress of children on their pathways towards literacy and numeracy to be tracked over time.

Initially called PIPS, *On-Entry Baseline Assessment* (Tymms et al., 2004), the instrument has the aim of providing schools with high quality data on student's progress in the first stages of their schooling. The data collected is processed and reports are generated and delivered individually to the teachers, concerning the conditions and progress of a class/group. They are presented at the individual level and contextualized with the advance of the class/group and their counterparts in other schools with similar characteristics (Tymms & Albone, 2002).

PIPS is composed of the following sections, which were all translated and adapted to the Brazilian context:

- a) Handwriting – the child is asked to write his/her own name;
- b) Vocabulary – the child is asked to identify objects embedded within a series of images;
- c) Ideas about reading – assessment of concepts of recognition of the written language;
- d) Phonological awareness – rhymes and repetitions;
- e) Letters Identification – a fixed order of mixed upper- and lower-case letters;
- f) Word recognition and reading – words, sentences, and comprehension;

- g) Ideas about mathematics— assessment of understanding of mathematical concepts (size, volume, etc.);
- h) Counting and ability to use numbers;
- i) Sums – addition and subtraction problems presented without symbols;
- j) Shape identification;
- k) Digit identification;
- l) Mathematical problems – including sums with symbols.

The application of the assessment is one-to-one and can be conducted by the teacher of the class/group or by another adult; the duration varies from 10 to 20 minutes. The computer program presents the questions orally, and, depending on the type of question, the child responds by pointing at the screen or by saying the answer. The teacher records the answer directly on the computer screen, and the program selects the next question. The program is adjusted to the child's responses. For each section, when there are three wrong answers in a row or four wrong altogether the program moves to the next section. Each section of the test presents items of progressive difficulty, which makes it possible to have a test with a minimal duration so that the child does not become bored with questions that are too simple or too difficult (Tymms et al., 2004).

iPIPS was derived from PIPS, with the addition of the term “International” to the title of the project⁶. Inspired by the success of the measurement program as a pedagogic tool, its providers, the Centre for Evaluation and Monitoring (CEM) at Durham University proposed its adoption in international comparative research.

At the present stage, iPIPS is being applied to children aged 4–7 in nine countries, having already generated several scientific works that attest to its technical quality and predictive power. The English version of PIPS has a high reliability level in the test/retest of 0.98 (CEM Centre, 1999) and good predictive power for reading and mathematics. Using a base measurement with children aged 4 or 5, the tests presented a correlation of 0.71 for reading three years after the first measurement, and a correlation of 0.7 after seven years (in England, the children were aged

⁶ See <<http://www.cem.org/ipips>>

11) (Tymms, 1999b; Tymms et al., 1997; Tymms et al., 2004). These are high values compared with those obtained in the other tests available, which serves to reinforce PIPS relevance for use in school effectiveness studies in longitudinal or experimental designs.

Performance Indicators in Primary Schools (PIPS) assessment: adaptation and pertinence for the Brazilian context

The PIPS assessment for language and mathematics was adapted and pre-tested in Brazil, so that it could be used in longitudinal studies. The study aims to follow children for three years, from preschool up to the first year of primary school (age range 4-7), in order to identify the characteristics of the institutions (school inputs, organization of the provision) and pedagogic processes associated to students' learning, in particular literacy and numeracy. The main collection tool for information about the state of children's development is the *Performance Indicators in Primary Schools*⁷. The longitudinal study was approved by the Ethics Committee (*Plataforma Brasil*) in 2016 (Document 1.625.525)

In 2016, four important steps in the preparation of the longitudinal study were carried out: the adaptation of the test (items) to Portuguese; the cultural adaptation of the figures presented in the test (application booklet); the production of the software that aid researchers in the application of the test; and the pre-test of the items in a sample of students from public and private schools. The adaptation of the instrument to Portuguese was performed based on the following protocol: (i) four independent translators worked on the mathematics and Portuguese tests (two on each); (ii) remaining doubts or discrepancies in the translations were debated and, when there was no consensus, an independent translator was called to assist with the final decision; (iii) a reverse translation of the material, from Portuguese back to English, by a translator who had not participated in the first adaptation phase of the instrument.

The pre-test was applied in a total of 560 children in three cities: Rio de Janeiro, Juiz de Fora, and Petrolina. Table 1 presents the distribution of approaches taken by the researchers.

⁷ Besides the aforementioned instruments, the longitudinal study also collected contextual data on the students (socio-demographic characteristics) and the schools—including both the characteristics of the provision (school inputs, organization of the provision, teacher training) and the pedagogic processes, focusing on those related with literacy and numeracy (lesson planning and organization, internal organization of groups/classes, classroom atmosphere, interaction between families and schools).

Table 1 – Application of the iPIPS Brazil baseline pre-test

City	Use of the software	Total applications
Rio de Janeiro municipal public system	Yes	399
Rio de Janeiro private school	No	17
Juiz de Fora municipal public system	Yes	92
Petrolina municipal public system	No	52

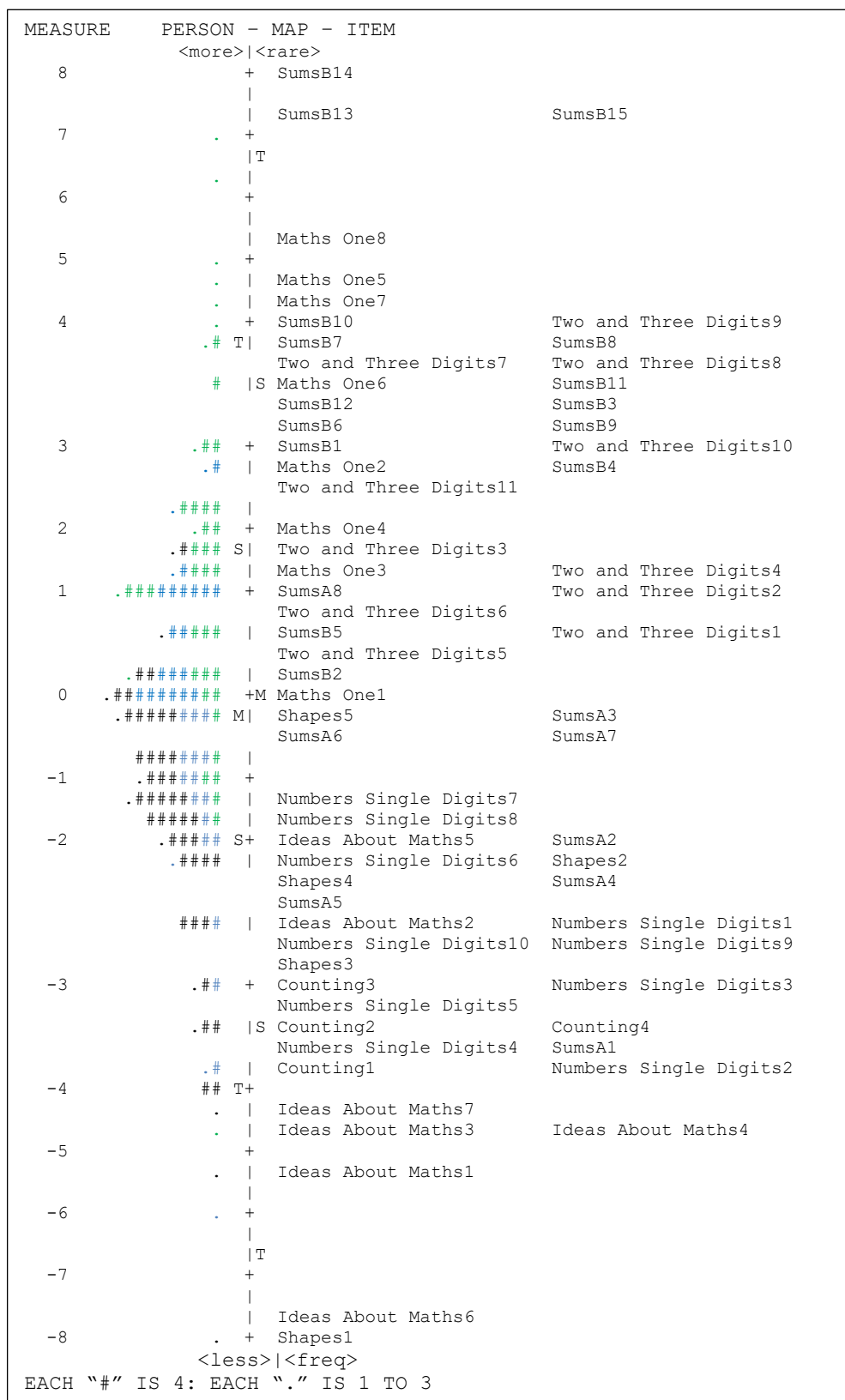
Source: LaPOpe UFRJ

In the pre-test, the PIPS instruments were applied to children aged 4–7, in groups with a quite heterogeneous home background profile. The main aim was to test the suitability of the assessment and spot possible limitations in translation or in the application protocol. Three main analyses were made with the data collected in the pre-test using Rasch measurement (Bond & Fox, 2015; Boone, 2006) with Winstep software: This focused on (i) the behavior of the items of language and mathematics; (ii) the distribution of the items and students on the scale—Wright Map; (iii) the correlation of the individual measurements of the students in the iPIPS with contextual variables.

The Wright Map below shows the results for mathematics considering children enrolled in public schools of two cities⁸. The Wright Map produced by Winstep shows the distribution of 491 children along the scale (left hand side of the map; each “#” represents 4 children, each “.” represents 3 or less children), the distribution of the items along the scale (right hand side of the map) shows the difficulty level of the items. The probability of getting a correct response from a child for any item measure can be calculated using the difference between the two on the vertical scale. Children are shown in three different colors depending on the age/group: (i) Black – 1st year of preschool, (ii) Blue – 2nd year of preschool, (iii) Green – 1st year of primary school.

⁸ The pre-test involved children enrolled the first and second years of preschool, as well as students enrolled in the first year of primary school, distributed in 13 schools.

Figure 1 – Wright Map of the mathematics pre-test on municipal public schools – 491 children assessed



Source: LaPOpe UFRJ

The preliminary analyses indicate that the items of both tests present adequate behavior, suggesting the theoretical adequacy of the items and a good adaptation and application protocol. Analyzing only mathematics items, Person Reliability is .92 and Item Reliability .97. The map of the items generated by Winstep suggests that the PIPS scale is adequate for Brazilian students aged 4–7.

The distribution of the data (children) for each group is very close to a normal curve and the items are well distributed throughout the scale. More importantly, the level of difficulty of the different sections corresponds to the theoretical presupposition used in devising the test. For example, items about identification of letters are, on average, easier to get right than reading words or short sentences. The same was observed regarding language: items involving identifying letters were easier compared with items that demanded reading words or passages.

The ladder (Tymms, Howie, Merrell, Combrinck, & Copping, 2017) in Table 2 presents the overview of the mathematics results and its pedagogical interpretation for the three grades (first and second years of preschool and first year of primary education) evaluated in 13 schools during the pre-test:

Table 2 – Math ladder

				First Year Preschool	Second Year Preschool	First Year Primary School
Advanced		Able to do harder formal sums		0,0%	1,4%	5,0%
		Able to do complex mental arithmetic				
		Able to count coins				
		Fourth rung				
Simple formal arithmetic		Identifies three-digit numbers		1,2%	6,1%	24,6%
		Able to do formal sums				
		Able to identify coins				
		Able to count on				
		Able to use simple fractions				
		Third rung				
Simple formal arithmetic		Identifies two-digit numbers		21,6%	41,4%	54,7%
		Able to do harder informal sums				
		Can do simple formal sums				

		Second rung				
Informal arithmetic		Identifies numbers 6 to 10		65,4%	42,9%	15,1%
		Able to do simple informal sums				
		Knows words such as “most” and “more”.				
		First rung				
Ground level		Identifies numbers 1 to 5		11,7%	8,2%	0,6%
		Counting a few objects by rote				
		Knows what “how many” means				

Source: LaPOpe UFRJ

As expected, Table 2 shows a different distribution of students enrolled at the first and second years of preschool and first year of primary school along the rungs of the ladder. For example, at the end of the first year of preschool, most children evaluated were at the informal arithmetic level. They were able to identify single-digit numbers and were able to do simple informal sums. However, most children evaluated at the end of the first year of elementary education were at the simple formal arithmetic level. They had developed abilities such as identifying two-digit numbers and doing harder informal sums and simple formal sums. Small isolated problems in the Application Booklet (support material offered to the students at the moment of the test) and in the application protocol were identified by the team and were corrected before the application of the first wave of the longitudinal study in March 2017.

The final analyses were undertaken using the data collected in the city of Rio de Janeiro, especially for pupils enrolled on the first grade of primary school. For this particular age group, we had further data available such as scores in SME-RJ standardized tests and family background information/variables. The PIPS test was applied in six primary schools and in each school only one classroom was assessed. The variables used in the analyses are described in Chart 1 and Table 3 below:

Chart 1 – Variables used in the models

Variable	Type	Description
Dependent Variables		
Mathematics score (PIPS)	Continuous	Mathematics score estimated using PIPS instrument
Language score (PIPS)	Continuous	Language score estimated using PIPS instrument
Mathematics score (Rio)	Continuous	Mathematics score in Rio de Janeiro's Local Assessment
Language score (Rio)	Continuous	Language score in Rio de Janeiro's Local Assessment
Independent variables		
Color White	Dummy	Indicate if Pupil is white (1=yes/0=no)
Conditional cash transfer	Dummy	Indicate if Pupil's family/household receives conditional cash transfer (1=yes/0=no)
Sex – Boy	Dummy	Indicates if Pupil is a boy (1=yes/0=no)
Age	Numeric	Pupil's age (months)
Parental Education		
Elementary education	Dummy	Indicates the parents have at most complete elementary education (1=yes/0=no)
Secondary education	Dummy	Indicates that at least one of the parents completed secondary education (1=yes/0=no)
Higher Education	Dummy	Indicates that at least one of the parents completed higher education (1=yes/0=no)

Source: LaPOpe UFRJ

Table 3 – Descriptive statistics of the variables used in the models

	Obs	Min	Max	Mean	Std Dev
Mathematics score (IPIPS)	140	-2,82	6,27	1,00	1,61
Language score (IPIPS)	140	-2,74	4,42	0,93	1,17
Mathematics score (Rio)	123	83,66	236,52	189,17	39,99
Language score (Rio)	126	51,23	237,41	189,61	38,10
Color White	133	0	1	0,444	0,499
Conditional cash transfer	137	0	1	0,256	0,438
Sex – Boy	140	0	1	0,486	0,502
Age					
Parental Education	119	%			
Elementary Education		51,26			
Secondary Education		39,50			
Higher Education		9,24			

Source: LaPOpe UFRJ

First, we note the correlation between the pupils' language and mathematics scores in the PIPS and the SME-RJ tests:

Table 4 – Correlation between pupil's language and mathematics scores – PIPS and SME-Rio*

		Language (PIPS)	Mathematics (PIPS)	Language (Rio)	Mathematics (Rio)
Language (PIPS)	Pearson correlation	1	,719**	,422**	,423**
	Sig.		,000	,000	,000
	N	118	118	118	118
Mathematics (PIPS)	Pearson correlation	,719**	1	,329**	,471**
	Sig.	,000		,000	,000
	N	118	118	118	118
Language (Rio)	Pearson correlation	,422**	,329**	1	,597**
	Sig.	,000	,000		,000
	N	118	118	118	118
Mathematics (Rio)	Pearson correlation	,423**	,471**	,597**	1
	Sig.	,000	,000	,000	
	N	118	118	118	118

Note: two-tailed test.

*Only the pupils with the four measures were considered in the analyses.

Source: LaPOpe UFRJ

Table 4 shows a strong correlation between pupils' language and mathematics scores using PIPS test ($r=0,719$). Considering SME standardized tests, we have found a slightly weaker correlation ($r=0,597$).

The analyses also included linear regression models estimating two outcomes: pupils' individual mathematics and language scores for PIPS and the SME-RJ standardized test using contextual information about the families as independent variables (information obtained in the *Sistema de Gestão Acadêmica – SGA*):

Table 5 – Linear regression models estimating language scores

	Language (Rio)	Language (PIPS)
	(1)	(1)
Constant	154,834 [62,755]**	-6,467 [1,869]***
Parental Education		
Secondary Education	8,028 [7,151]	0,335 [0,213]
Higher Education	30,894 [11,961]**	1,534 [0,356] ***
Color (white)	-8,397 [6,773]	-0,354 [0,202] *
Conditional Cash Transfer	-5,879 [7,437]	-0,262 [0,222]
Sex	-3,584 [6,694]	-0,104 [0,199]
Age	0,456 [0,723]	0,088 [0,022]***
N	103	103
R2	0,089	0,279

* significant at 10%, ** significant at 5%, *** significant at 1%

Standard errors in brackets.

Only the students that had scores for both language tests were considered.

Source: LaPOpe UFRJ

Table 6 – Linear regression models estimating mathematics scores

	Mathematics (Rio)	Mathematics (PIPS)
	(1)	(1)
Constant	107,788 [77,677]	-9,649 [2,907]***
Parental Education		
Secondary Education	18,539 [7,984]**	0,564 [0,299]*
Higher Education	41,618 [14,489] ***	2,572 [0,542]***
Color (white)	-7,24 [7,686]	-0,402 [0,288]
Conditional Cash Transfer	-7,039 [8,374]	-0,016 [0,313]
Sex	-16,097 [7,503]**	-0,136 [0,281]
Age	1,004 [0,897]	0,122 [0,034]***
N	101	101
R2	0,166	0,261

* significant at 10%, ** significant at 5%, *** significant at 1%

Standard errors in brackets.

Only the students that had scores for both math tests were considered.

Source: LaPOpe UFRJ

The linear regression model, with no school fixed effects, suggests that 27,9% and 26,1% of the variance of the language and mathematics scores using the PIPS test is explained by the background variables. The analyses performed with standardized tests applied by the SME-RJ (*Alfabetiza Rio* – 1st grade of Primary School) to the same students indicate that the same contextual variables explain 8,9% and 16,6% of the variance of the language and mathematics scores, respectively. The results strengthen the hypothesis that the measurements generated by PIPS are of higher quality at the individual level and correspond to the results for Brazilian students and to other international research suggesting that the socio-economic level explains, on average, 30% of the variance in the standardized tests in the USA (Sirin, 2005).

Conclusion

Measuring and monitoring children's development is a key aspect of assessing the impact of educational policy and producing reliable measures about the quality of the educational system and outcomes. In Brazil, as in other countries, it is possible to observe resistance from some educational researchers regarding the potential harm that measuring children could produce. We believe that we need to develop measures in order to enhance knowledge about what children know and can do when they start school and how much progress is made in the first three years in school. A longitudinal design allows researchers to identify the most effective approaches to help all children, including those from disadvantaged backgrounds. Not knowing how children are progressing could harm poor and disadvantaged children and increase school inequality.

We have argued for the importance of having early baseline measures for children development in order to construct robust research designs that allow causal inference—specially as the existing large-scale assessment instruments in Brazil are not adequate for such studies due to their cross-sectional nature and the absence of a baseline measure. Therefore, iPIPS can be a first step for the use of longitudinal studies with this purpose as the pre-test results showed its pertinence for the Brazilian context.

The preliminary analyses indicate that the items of both tests, mathematics and language, present adequate behavior, suggesting the theoretical adequacy of the items and a good adaptation and application protocol. Analyzing only mathematics items, Person Reliability is .92 and Item Reliability is .97. The map of the items generated by Winstep suggests that the PIPS scale is adequate for Brazilian students aged 4–7.

Measurements generated by PIPS are of higher quality at the individual level and correspond to the results for Brazilian students and to other international research suggesting that the socio-economic level explains, on average, 30% of the variance in the standardized tests in the USA (Sirin, 2005). The distribution of the data (children) for each group is very close to a normal curve and the items are well distributed throughout the scale. More importantly, the level of difficulty of the different sections corresponds to the theoretical presupposition used in devising the test.

These kinds of studies can provide information for taking decisions about future educational policies for the pre-school and the first year of primary school. We consider that there is an irreversible trend, both in Brazil and in the rest of the world, towards the adoption of systematic information, collected with growing rigor and precision, as the basis for educational policy decision-making—the so-called evidence-based policies (Dagenais et al., 2012; Miller & Pasley, 2012; Nutley, Morton, Jung, & Boaz, 2010).

References

- Bonamino, A. M. C., & Oliveira, L. H. G. (2013). Estudos longitudinais e pesquisa na educação básica. *Linhas Críticas*, 19(38), 33-50. doi:10.26512/lc.v19i38.4084
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: fundamental measurement in the human sciences*. New York: Routledge.
- Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how? *CBE Life Sciences Education*, 15(4). doi:10.1187/cbe.16-04-0148
- Brasil. (1988). *Constituição Federal do Brasil*. Brasília: Senado.
- Brasil. (2006). Ministério da Educação. Secretaria de Educação Básica. *Parâmetros nacionais de qualidade para a educação infantil* (Vol. 1). Brasília: Ministério da Educação.
- Brasil. (2015). *Plano Nacional de Educação PNE 2014-2024: linha de base*. Brasília: INEP.
- Brasil. (2018, 28 de junho). Ministério da Educação. *Ministro da Educação anuncia mudanças no sistema de avaliação a partir de 2019*. Retrieved from <http://portal.mec.gov.br/ultimas-noticias/211-218175739/65791-ministro-da-educacao-anuncia-mudancas-no-sistema-de-avaliacao-a-partir-de-2019>
- Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001) The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. *Developmental Psychology*, 37(2), 231-242.
- Campos, M. M. (Coord.). (2010). *Educação infantil no Brasil: avaliação qualitativa e quantitativa*. São Paulo: Fundação Carlos Chagas.

- Campos, M. M., Bhering, E. B., Esposito, Y., Gimenes, N., Abuchaim, B., Valle, R., & Unbehaum, S. (2011). A contribuição da educação infantil de qualidade e seus impactos no início do ensino fundamental. *Educação e Pesquisa*, 37(1), 15-33.
- CEM Centre. (1999). Performance Indicators in Primary Schools: Baseline Assessment: Technical Report [CD-ROM Version]. Durham: University of Durham.
- Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on a life cycle skill formation. In E. A. Hanushek, & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 1, pp. 697-812). Amsterdam: Elsevier.
- Dagenais, C., Lysenko, L., Abrami, P. C., Bernand, R. M., Ramde, J., & Janosz, M. (2012). Use of research-based information by school practitioners and determinants of use: a review of empirical research. *Evidence & Policy: A Journal of Research, Debate and Practice*, 8(3), 285-309.
- Damiani, M. F., Dumith, S., Horta, B. L., & Gigante, D. (2011). Educação infantil e longevidade escolar: dados de um estudo longitudinal. *Estudos em Avaliação Educacional*, 22(50), 515-532. doi:10.18222/cae225020111968
- Franco, C. (2001). O SAEB: Sistema de Avaliação da Educação Básica: potencialidades, problemas e desafios. *Revista Brasileira de Educação*, (17), 127-155. doi:10.1590/S1413-24782001000200010
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, 54(1), 3-56.
- Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry*, 46(3), 289-324.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (2014). *Censo escolar da educação básica 2013: resumo técnico*. Brasília: INEP.
- Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, 20(2), 82-88.
- Lei nº 9.394, de 20 de dezembro de 1996. (1996, 23 de dezembro). Estabelece as diretrizes e bases da educação nacional. *Diário Oficial [da República Federativa do Brasil]*, v. 134, n. 248.

- Miller, B., & Pasley, J. (2012). What do we know and how well do we know it? Identifying practice-based insights in education. *Evidence & Policy: A Journal of Research, Debate and Practice*, 8(2), 193-212. doi:10.1332/174426412X640081
- Nutley, S., Morton, S., Jung, T., & Boaz, A. (2010). Evidence and policy in six European countries: diverse approaches and common challenges. *Evidence and Policy: A Journal of Research, Debate and Practice*, 6(2), 131-144. doi:10.1332/174426410X502275
- Observatório do PNE. (n.d.). Retrieved from <https://www.observatoriodopne.org.br/conteudos/analises>
- Peisner-Feinberg, E., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., ... Zelazo, J. (2000). *The children of the cost, quality, and outcomes study go to school: technical report*. Chapel Hill: University of North Carolina at Chapel Hill.
- Peisner-Feinberg, E., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72(5), 1534-1553.
- Rosemberg, F. (1999a). Expansão da educação infantil e processos de exclusão. *Cadernos de Pesquisa*, (107), 7-40. doi:10.1590/S0100-15741999000200001
- Rosemberg, F. (1999b). O estado dos dados para avaliar políticas de educação infantil. *Estudos em Avaliação Educacional*, (20), 5-58. doi:10.18222/ae02019992234
- Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Hunt, S. (2008). *Influences on children's attainment and progress in key stage 2: cognitive outcomes in year 6*. Nottingham: Department for Children, Schools and Families.
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J., & Smedt B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, 20(3), e12372. doi:10.1111/desc.12372
- Schweinhart, L. J. (2002). How the High/Scope Perry Pre-school Study grew: a researcher's tale. *Phi Kappa Delta Research Bulletin*, 32.

- Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.
- Sylva, K., Melhuish, E., & Sammons, P. (Eds.). (2010). *Early childhood matters: evidence from the effective pre-school and primary education project*. London: Routledge.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., Taggart, B., & Elliot, K. (2003). *The effective provision of pre-school education (EPPE) project: findings from the pre-school period*. London: University of London.
- Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly*, 21(1), 76-92.
- Tymms, P. (1999a). *Baseline assessment and monitoring in primary schools: achievements, attitudes and value-added indicators*. London: David Fulton.
- Tymms, P. (1999b). Baseline assessment, value-added and the prediction of reading. *Journal of Research in Reading*, 22(1), 27-36.
- Tymms, P., & Albane, S. (2002). Performance indicators in primary schools. In A. J. Visscher, & R. Coe (Eds.), *School improvement through performance feedback* (pp. 191-218). London: Routledge.
- Tymms, P., Howie, S., Merrell, C., Combrinck, C., & Copping, L. (2017). *The first year at school in the western cape: growth, development and progress*. London: Nuffield Foundation.
- Tymms, P., Jones, P. R., Albane, S. A., & Henderson, B. E. (2009). The first seven years at school. *Educational Assessment, Evaluation and Accountability*, 21, 67-80. doi:10.1007/s11092-008-9066-7
- Tymms, P., Merrell, C., & Henderson, B. (1997). The first year at school: a quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation*, 3(2), 101-118.
- Tymms, P., Merrell, C., & Jones, P. (2004). Using baseline assessment data to make international comparisons. *British Educational Research Journal*, 30(5), 673-689. doi:10.1080/0141192042000234647

Consulted references

- Brasil. (2006, 20 de dezembro). Emenda Constitucional nº 53, de 19 de dezembro de 2006. Dá nova redação aos arts. 7º, 23, 30, 206, 208, 211 e 212 da Constituição Federal e ao art. 60 do Ato das Disposições Constitucionais Transitórias. *Diário Oficial da União*.
- Franco, C., Brooke, N., & Alves, F. (2008). Estudo longitudinal sobre qualidade e equidade no ensino fundamental brasileiro: GERES 2005. *Ensaio: Avaliação e Políticas Públicas em Educação*, 16(61), 625-637. doi:10.1590/S0104-40362008000400008
- Lima, A. L., & Huaira-Contreras, C. A. (2012). Evolução do analfabetismo funcional no Brasil: 2001-2010. *Estudos em Avaliação Educacional*, 23(52), 106-128.
- Lei nº 12.796, de 4 de abril de 2013. (2013, 5 de abril). Altera a Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação nacional, para dispor sobre a formação dos profissionais da educação e dar outras providências. *Diário Oficial da União*.

Submission data: Submitted to evaluation on April 27, 2018; revised on November 28, 2018; accepted for publication on February 24, 2019.

Corresponding author: Tiago Lisboa Bartholo - Universidade Federal do Rio de Janeiro, Faculdade de Educação, Avenida Pasteur, 250 (fundos), Urca, Rio de Janeiro, RJ, Brazil, 22290-902